Time-varying mixture GARCH models and asymmetric volatility

Haas, Markus ; Krause, Jochen ; Paolella, Marc S ; Steude, Sven C

Abstract: The class of mixed normal conditional heteroskedastic (MixN-GARCH) models, which couples a mixed normal distributional structure with GARCH-type dynamics, has been shown to offer a plausible decomposition of the contributions to volatility, as well as excellent out-of-sample forecasting performance, for financial asset returns. In this paper, we generalize the MixN-GARCH model by relaxing the assumption of constant mixing weights. Two different specifications with time–varying mixing weights are considered. In particular, by relating current weights to past returns and realized (component-wise) likelihood values, an empirically reasonable representation of Engle and Ng's (1993) news impact curve with an asymmetric impact of unexpected return shocks on future volatility is obtained. An empirical out-of-sample study confirms the usefulness of the new approach and gives evidence that the leverage effect in financial returns data is closely connected, in a non-linear fashion, to the time–varying interplay of mixture components representing, for example, various groups of market participants.

DOI: https://doi.org/10.1016/j.najef.2013.02.024

Originally published at:
Haas, Markus; Krause, Jochen; Paolella, Marc S; Steude, Sven C (2013). Time-varying mixture GARCH models and asymmetric volatility. North American Journal of Economics and Finance, 26:602-623.
DOI: https://doi.org/10.1016/j.najef.2013.02.024

Time-varying Mixture GARCH Models
and
Asymmetric Volatility*

Markus Haas[a] Jochen Krause[b,†] Marc S. Paolella[b,c] Sven C. Steude[b]

[a] *Institute for Quantitative Business and Economics Research, University of Kiel, Germany*

[b] *Department of Banking and Finance, University of Zurich, Switzerland*

[c] *Swiss Finance Institute*

January 25, 2013

Abstract

The class of mixed normal conditional heteroskedastic (MixN-GARCH) models, which couples a mixed normal distributional structure with GARCH-type dynamics, has been shown to offer a plausible decomposition of the contributions to volatility, as well as excellent out-of-sample forecasting performance, for financial asset returns. In this paper, we generalize the MixN-GARCH model by relaxing the assumption of constant mixing weights. Two different specifications with time–varying mixing weights are considered. In particular, by relating current weights to past returns and realized (component-wise) likelihood values, an empirically reasonable representation of Engle and Ng's (1993) news impact curve with an asymmetric impact of unexpected return shocks on future volatility is obtained. An empirical out-of-sample study confirms the usefulness of the new approach and gives evidence that the leverage effect in financial returns data is closely connected, in a non-linear fashion, to the time–varying interplay of mixture components representing, for example, various groups of market participants.

Keywords: GARCH; News Impact Curve; Leverage Effect; Down-Market Effect; Mixtures; Time–Varying Weights; Value–at–Risk

JEL classification: C22; C51; G10

*This paper is a highly altered and extended version of the manuscript Haas et al. (2006b). The authors are grateful to Emese Lazar, Michael McAleer, and an anonymous referee, for discussions and comments which led to a substantially improved paper. The research of Haas was supported by the *Deutsche Forschungsgemeinschaft*. Part of the research of Paolella and Steude has been carried out within the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK), which is a research program supported by the Swiss National Science Foundation.

†Corresponding author: `jochen.krause@bf.uzh.ch`

1 Introduction

Among the many apparent empirical regularities of financial time series, one of the most researched is the relationship between equity returns and volatility. Over the last two decades, a large amount of literature reports an asymmetric volatility response between positive and negative returns.

The initial discovery of asymmetry in the relationship between returns and volatility is usually accredited to Black (1976) and Christie (1982) with their observation that current returns and future volatility are negatively correlated, commonly referred to in the literature as Black's leverage effect. The historic explanations of such market behavior are grounded in the firms' dept–equity ratio that changes with movements in the return and, thus, alters the stock's riskiness. However, an increasing number of studies challenge this fundamental reasoning. For example, Hasanhodzic and Lo (2011) find the leverage effect also present in all–equity financed companies and report its effect even stronger than for leveraged firms. Similarly, Hens and Steude (2009) find the effect in the laboratory environment absent of any leverage implying that the inverse relationship between price and volatility is not driven by financial leverage. In addition, Figlewski and Wang (2000) present evidence that the leverage effect is largely independent of a change in the firms' capital structure. More evidence for the leverage effect in assets for which the traditional explanation cannot hold is provided in Park (2011), who conjectures a herding type of behavior to explain it.

Insight into the asymmetric volatility response has also risen from a different strand of literature. The ARCH and GARCH model classes – that in their original version are such that negative and positive return shocks have the *same* impact on volatilities – have been extended by several authors to include asymmetric effects as well, e.g., in the univariate case, models that allow for this effect include the EGARCH model of Nelson (1991), the GJR model of Glosten et al. (1993) and the threshold ARCH model of Zakoian (1994). All of these volatility models are asymmetric in a sense that "bad" news tend to be associated with a larger increase in (tomorrow's) volatility than "good" news of the same magnitude – but positive news does not *reduce* volatility like the leverage effect suggests; compare Asai and McAleer (2011). Only the EGARCH model, although not guaranteeing the leverage effect, permits the effect subject to restrictions on the size and sign parameters.

Some elements of the structure in the return-volatility relationship are not fully understood yet and there is still a lively debate about its dynamics. In this paper, we propose a mixture GARCH approach that can represent a variety of different asymmetric response patterns. The model yields a new and flexible dynamic structure for modelling the (generally) asymmetric relationship between returns and volatility that allows feedback between different components of variances and the overall process. The goal of this paper is to study these volatility dynamics in detail. Our proposed model has a rich GARCH structure, so that an increase in volatility occurs when a negative or positive shock hits the market, but its impact is enhanced for negative shocks, while mitigated for positive shocks.

Further, the use of a mixed normal distribution for modeling the unconditional distribution of asset returns has been considered by numerous authors, including Fama (1965), Kon (1984), Tucker and Pond (1988), and Aparicio and Estrada (2001). More recently, Kim and White (2004, p. 72) provide further evidence of the appropriateness of normal mixtures for financial data, stating "[We propose that] it may be more productive to think of the S&P500 index returns studied here as being better described as a mixture containing a predominant component that is nearly symmetric with mild kurtosis and a relatively rare component that generates highly extreme behavior." Along similar lines, Neftci (2000) argues that the extreme movements in asset prices are caused by mechanisms which are "structurally different" from the "routine functioning of markets".

The problem with any unconditional model for asset returns is that they cannot capture the blatant volatility clustering inherent in virtually all return series observed at weekly or higher frequencies, and will suffer appropriately in terms of short–term Value–at–Risk (VaR) forecasting ability. The effectiveness and easy implementation of GARCH models for this purpose is undisputed, and numerous variations and extensions of Bollerslev's (1986) original construct have been proposed and shown to deliver superior forecasts; see, for example, Palm (1996), Kuester et al. (2006), and Alexander (2008, Ch. 4) for surveys.

The mixed normal GARCH, or MixN-GARCH, is a relatively recent GARCH–type model class which combines the features of normal mixture distributions and a GARCH model, and has been independently proposed and investigated by Alexander and Lazar (2006) and Haas et al. (2004a,b). By judiciously coupling a k–mixture of normal distributions with a GARCH–type dynamic structure that links the k density components, several previously advocated models are nested, and a variety of stylized facts of asset returns can be successfully modeled, such as the usual fat tails and volatility clustering, but also time–varying skewness and kurtosis. The model has been shown in the aforementioned papers to offer a plausible decomposition of the contributions to market volatility, and also to deliver highly competitive out-of-sample forecasts. For further detail and more recent extensions, see Haas and Paolella (2012).

A common property of MixN-GARCH is the constancy of mixing weights of the component densities, which often allows for a straightforward interpretation of the impact of the individual components. However, constancy of the distributional proportions is not necessarily a realistic assumption, and, as we demonstrate below, leads to less accurate forecasts compared with a more general class of models which does allow for time–varying mixture weights.

While the use of mixtures, in particular, the mixture of normals distribution, is ubiquitous in numerous scientific applications, of which finance is only one of many examples (see, e.g., McLachlan and Peel, 2000; and Frühwirth-Schnatter, 2010), the thorny, and very real, issue of avoiding the singularities when maximizing the likelihood needs to be addressed. We employ the new, easily implemented, and theoretically very attractive method introduced in Broda et al. (2013), which is applicable to unconditional mixtures, as well as mixture-GARCH models. This renders model estimation to be very simple to implement, as fast as standard likelihood optimization, numerically fully unproblematic, and, under appropriate conditions on the data generation process, the resulting maximum likelihood estimates are consistent.

Anticipating the empirical results in Section 4, the newly proposed model, denoted MixN-GARCH-LIK, performs very well according to numerous out-of-sample criteria for the majority of the considered data (seven major equity indices and exchange rates). Given the ease of use in implementation and estimation, as well as the general appeal of mixture distributions in finance, from both economic and empirical perspectives, we show that the new class of models provide a worthy contribution for forecasting the distribution and tail risk of univariate financial returns data.

The remainder of this paper is as follows. Section 2 briefly introduces the MixN-GARCH model. Section 3 discusses its extension to allow for time–varying mixing weights and reviews the implied news impact curves. Section 4 details an empirical exercise, and Section 5 concludes.

2 Mixed Normal GARCH

In the mixed normal GARCH (MixN-GARCH) model the conditional density of return r_t is assumed to be a *finite normal mixture distribution* with k components. That is, with f_t denoting a conditional density based on the information set at time t,

$$f_{t-1}(r_t; \lambda_{1t}, \ldots, \lambda_{k,t}, \mu_{1t}, \ldots, \mu_{k,t}, \sigma^2_{1t}, \ldots, \sigma^2_{k,t}) = \sum_{j=1}^{k} \lambda_{jt} \phi(r_t; \mu_{jt}, \sigma^2_{jt}), \tag{1}$$

where

$$\phi(r_t; \mu_{jt}, \sigma^2_{jt}) = \frac{1}{\sqrt{2\pi}\sigma_{jt}} \exp\left\{ -\frac{(r_t - \mu_{jt})^2}{2\sigma^2_{jt}} \right\}$$

is the normal density, the strictly positive *mixing weights* (or *probabilities*) λ_{jt} satisfy $\sum_j \lambda_{jt} = 1$, and the $k \times 1$ vector $\boldsymbol{\sigma}_t^{(2)} = (\sigma_{1t}^2, \ldots, \sigma_{k,t}^2)'$ of conditional *component variances* follows a GARCH(p, q) process of the form

$$\boldsymbol{\sigma}_t^{(2)} = \boldsymbol{\omega} + \sum_{i=1}^{q} \boldsymbol{\alpha}_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \boldsymbol{\beta}_i \boldsymbol{\sigma}_{t-i}^{(2)}, \tag{2}$$

where the error term

$$\varepsilon_t = r_t - \mathbb{E}_{t-1}[r_t] = r_t - \sum_{j=1}^{k} \lambda_{jt} \mu_{jt}, \tag{3}$$

and $\boldsymbol{\omega} \in \mathbb{R}^k$, $\boldsymbol{\alpha}_i \in \mathbb{R}^k$, $i = 1, \ldots, q$, and $\boldsymbol{\beta}_i \in \mathbb{R}^{k \times k}$, $i = 1, \ldots, p$, are parameters matrices which have to obey restrictions to guarantee that $\boldsymbol{\sigma}_t^{(2)}$ remains positive for all t. As with the standard (single-component) GARCH model, $p = q = 1$ is typically found to be sufficient; moreover, the *diagonal* GARCH specification with a diagonal $\boldsymbol{\beta}_1$ is typically favored in empirical applications and admits a clear-cut interpretation of the component-specific volatility processes (see Haas et al., 2004b; and Haas and Paolella, 2012, Section 3.2.3, for discussion). In this case, we write the model as

$$\boldsymbol{\sigma}_t^{(2)} = \boldsymbol{\omega} + \boldsymbol{\alpha} \varepsilon_{t-1}^2 + \boldsymbol{\beta} \boldsymbol{\sigma}_{t-1}^{(2)}, \tag{4}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_k)' > \mathbf{0}$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)' \geq \mathbf{0}$, and $\boldsymbol{\beta} = \mathrm{diag}(\beta_1, \ldots, \beta_k) \geq \mathbf{0}$, where the inequalities have to hold element-wise.

The conditional mean of r_t has already been introduced in (3). Its conditional variance implied by the mixture density (1) is of great interest in the discussion that follows and is given by

$$\mathbb{V}_{t-1}(r_t) = \sum_{j=1}^{k} \lambda_{jt} (\sigma_{jt}^2 + \mu_{jt}^2) - \left(\sum_{j=1}^{k} \lambda_{jt} \mu_{jt} \right)^2. \tag{5}$$

Alexander and Lazar (2006) and Haas et al. (2004b) consider the case where the mixing weights, λ_{jt}, and the component means, μ_{jt}, are constant over time, but the generalization in Equations (1)–(3), with these quantities being time–varying, is conceptually straightforward. In this paper, we consider MixN-GARCH specifications with time–varying weights to capture an asymmetric impact of negative and positive and/or small and large shocks on future volatility, as discussed in the introduction.

3 Time-varying weights

The idea of modeling economic variables using mixtures with time–varying mixing weights (or regime probabilities) is not new. Most notably perhaps, the Markov–switching model of Hamilton (1989), which has many applications in macroeconomics and finance, can be interpreted in this framework. In addition, in a number of applications, mixture models with mixing weights depending on lagged process values as well as exogenous variables have been employed quite successfully. An example is the modeling of exchange rate behavior in target zones, where a jump component reflects the probability of realignments, and the probability of a jump depends on interest differentials and, possibly, further explanatory variables incorporating market expectations (see, e.g., Vlaar and Palm, 1993; Bekaert and Gray, 1998; Neely, 1999; Klaster and Knot, 2002 and Haas et al., 2006a). Cheng et al. (2009) provide an application to national stock index returns, and Tashman and Frey (2009) successfully use such models to capture a nonlinear relation between hedge fund returns and various market risk factors. The conditional densities of such mixture models exhibit an enormous flexibility. For example, as illustrated by Haas et al. (2006a) in an application to the EMS crisis of 1992, the predictive density may become bimodal when the probability of a realignment as well as the expected jump size are sufficiently large;

see also Wong and Li (2001) for an example of a bimodal predictive density in an hydrological application.

In this paper, we consider two different approaches to specifying time–varying weights in mixture GARCH models. In the first specification, to be discussed in Section 3.1, we let the weights depend on the lagged shock in a logistic fashion. By doing so, an asymmetric response of future volatility to negative and positive shocks is introduced, which is a robust feature of many stock return series. In the second variant, presented in Section 3.2, we follow a different approach and determine the conditional weights of the mixture components by their respective most recent explanatory power, as measured by their lagged component-specific likelihood contributions. This also induces a certain degree of asymmetry in the volatility response pattern since, for stock returns, there is also a *contemporaneous* negative relation between return and volatility, i.e., the high–volatility component is also that with the smaller expected return. In the outer parts of the distribution, however, volatility effects dominate, and thus both models represent different (potential) aspects of return volatility dynamics.

3.1 Time–varying mixture GARCH with logistic mixing weights

A general approach is to relate the weights of the components to past innovations via logistic response functions. In particular, in a two–component model, to which we restrict attention in this paper,[1] the weight of the first component is given by

$$\lambda_t(\boldsymbol{x}_t) = \frac{\exp\{\boldsymbol{\gamma}'\boldsymbol{x}_t\}}{1+\exp\{\boldsymbol{\gamma}'\boldsymbol{x}_t\}}, \tag{6}$$

with $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_{p-1})'$ being a vector of parameters and $\boldsymbol{x}_t$ a vector of p predetermined variables, typically including a constant.

A mixture GARCH model in this vein was considered by Bauwens et al. (2006), who use $\boldsymbol{x}_t = (1, \varepsilon_{t-1}^2)'$, i.e.,

$$\lambda_t(\varepsilon_{t-1}) = \frac{\exp\{\gamma_0 + \gamma_1\varepsilon_{t-1}^2\}}{1+\exp\{\gamma_0 + \gamma_1\varepsilon_{t-1}^2\}}, \tag{7}$$

so that, if $\gamma_1 > 0$, $\lambda_t \to 1$ as ε_{t-1}^2 becomes large. The motivation for this specification is that "large shocks have the effect of relieving pressure by reducing the probability of a large shock in the next period".[2]

Note that, in (7), λ_t is a symmetric function of ε_{t-1}, and $\lambda_t(-\infty) = \lambda_t(\infty) = 1$. In this paper, we aim at modeling an asymmetric effect of past shocks, and thus we let $\boldsymbol{x}_t = (1, \varepsilon_{t-1})'$ in (6), i.e., we specify the conditional mixing weight as

$$\lambda_t(\varepsilon_{t-1}) = \frac{\exp\{\gamma_0 + \gamma_1\varepsilon_{t-1}\}}{1+\exp\{\gamma_0 + \gamma_1\varepsilon_{t-1}\}}, \tag{8}$$

which, when coupled with the MixN-GARCH structure described in Section 2, will be referred to as MixN-GARCH-LOG model. The logic behind (8) is as follows. Suppose that the first component is the high–volatility regime, and $\gamma_1 > 0$. Then $\mathrm{d}\lambda_t/\mathrm{d}\varepsilon_{t-1} = \gamma_1\lambda_t(1-\lambda_t) > 0$, and, in view of (5), the conditional variance will be lower for positive shocks than for negative shocks of the same magnitude. In the next section, the capability of the MixN-GARCH-LOG model to reproduce various asymmetric response patterns is further elucidated via the concept of the news impact curve, as introduced in Engle and Ng (1993).

[1] For possible generalizations to k components, see Haas et al. (2006b).

[2] This assumes that the second component represents the high–volatility regime. In an application to the NASDAQ index, the authors find that, when using (7), the evidence for a time–varying mixing weight is weak.

3.1.1 Special cases and relation to other models

In this section, we illustrate the flexibility of the MixN-GARCH-LOG model introduced in Sections 2 and 3.1 in capturing various asymmetric response patterns of the conditional volatility to previous shocks. We consider various simple special cases of the general model and relate those to some standard asymmetric GARCH models discussed in the literature.[3]

A convenient tool to characterize the impact of news on conditional volatility is the news impact curve (NIC) devised by Engle and Ng (1993). The NIC describes the relation between the conditional variance σ_t^2 and the lagged shock ε_{t-1}, with the lagged conditional variances in the GARCH recursion fixed at their unconditional values. As discussed by Engle and Ng (1993), the NIC of the GARCH is a quadratic function centered at $\varepsilon_{t-1} = 0$, whereas most asymmetric GARCH models have NICs which either still have their minimum at zero but with different slopes for positive and negative shocks or which admit asymmetries by centering the quadratic at a nonzero (usually positive) value. By imposing certain parameter restrictions and considering limiting cases of the MixN-GARCH-LOG model process, we can isolate these typical (and further) shapes of the NIC and thereby get a glimpse of the flexibility of the general (unrestricted) model. To discuss these restrictions, we reproduce the general (non-diagonal) MixN-GARCH(1,1)-LOG process for two components, i.e.,

$$\begin{pmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \end{pmatrix} = \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \varepsilon_{t-1}^2 + \begin{pmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1}^2 \\ \sigma_{2,t-1}^2 \end{pmatrix}, \tag{9}$$

and the weight of the first component is described by (8). To fully concentrate on the conditional volatility dynamics, we furthermore assume zero component means, i.e., $\mu_1 = \cdots = \mu_k = 0$.

Consider the situation where, in (9), $\beta_{1,2} = \beta_{2,1} = 0$ (diagonal model) and $\alpha_1 = \alpha_2 \equiv \alpha$ and $\beta_{1,1} = \beta_{2,2} \equiv \beta$, i.e., the intercepts differ, whereas the GARCH dynamics are the same in both components. To figure out the NIC for this specification, we observe that it is identical to the one suggested by Vlaar and Palm (1993), where $\sigma_{2t}^2 = \sigma_{1t}^2 + a$ for constant a. To see this, let L denote the lag operator and write the ARCH(∞) representation of σ_{2t}^2 as

$$\sigma_{2t}^2 = \frac{\omega_2}{1-\beta} + \frac{\alpha\varepsilon_{t-1}^2}{1-\beta L} = \frac{\omega_2 - \omega_1}{1-\beta} + \frac{\omega_1}{1-\beta} + \frac{\alpha\varepsilon_{t-1}^2}{1-\beta L} = a + \sigma_{1t}^2, \tag{10}$$

where $a = (\omega_2 - \omega_1)/(1-\beta)$. Therefore, from (5), the conditional variance is

$$\begin{aligned} \sigma_t^2 &= \lambda_t \sigma_{1t}^2 + (1-\lambda_t)\sigma_{2t}^2 = \lambda_t \sigma_{1t}^2 + (1-\lambda_t)(a + \sigma_{1t}^2) \\ &= (1-\lambda_t)a + \sigma_{1t}^2 = (1-\lambda_t)a + \omega_1 + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{1,t-1}^2. \end{aligned} \tag{11}$$

If $\gamma_1 > 0$ and $\omega_2 > \omega_1$, λ_t is increasing in ε_{t-1} and $a > 0$, and thus the NIC is asymmetric in that it is higher for negative than for positive shocks of the same magnitude. As in the AGARCH model of Engle (1990) its minimum is also located at a positive value, as can be seen by minimizing (11) with respect to ε_{t-1}, i.e., setting to zero

$$\frac{\partial \sigma_t^2}{\partial \varepsilon_{t-1}} = -\frac{\partial \lambda_t}{\partial \varepsilon_{t-1}} a + 2\alpha\varepsilon_{t-1} = -\gamma_1 \lambda_t (1-\lambda_t) a + 2\alpha\varepsilon_{t-1},$$

which, as long as $\lambda_t \in (0,1)$, and since $a > 0$ and $\gamma_1 > 0$, can only be zero for a positive ε_{t-1}. The unconditional expectation of σ_{1t}^2 in (11) is not known in closed form, but the NIC can still be drawn by evaluating it via simulation and an example is shown in Panel (a) of Figure 1.

If we further put $\gamma_0 = 0$ and consider the limiting case $\gamma_1 \to \infty$ in (8), so that

$$\lambda_t = \begin{cases} 1, & \text{if } \varepsilon_{t-1} > 0, \\ \frac{1}{2}, & \text{if } \varepsilon_{t-1} = 0, \\ 0, & \text{if } \varepsilon_{t-1} < 0, \end{cases} \tag{12}$$

[3] A recent investigation of various asymmetric GARCH specifications is Rodriguez and Ruiz (2012).

we obtain the sign-switching GARCH model of Fornari and Mele (1997), except that the right-hand side lagged variance in (11) is $\sigma^2_{1,t-1}$ rather than the overall variance σ^2_{t-1}. The NIC of this process has a somewhat unusual form; namely it is a "broken" parabola with the positive and negative arms having different intercepts. To calculate the NIC explicitly for this special case (i.e., obtain an explicit expression for $\mathbb{E}[\sigma^2_{1t}]$), we define an indicator variable $\mathbb{1}_t$ which is unity or zero, depending on whether ε_t is drawn from the first or second component, respectively. We can then write, with $\{\eta_t\}$ denoting an iid sequence of standard normal variables,

$$\begin{aligned}
\sigma^2_{1t} &= \omega_1 + \alpha\varepsilon^2_{t-1} + \beta\sigma^2_{1,t-1} \\
&= \omega_1 + \alpha\eta^2_{t-1}\{\mathbb{1}_{t-1}\sigma^2_{1,t-1} + (1-\mathbb{1}_{t-1})\sigma^2_{2,t-1}\} + \beta\sigma^2_{1,t-1} \\
&= \omega_1 + \alpha\eta^2_{t-1}\{\mathbb{1}_{t-1}\sigma^2_{1,t-1} + (1-\mathbb{1}_{t-1})(\sigma^2_{1,t-1} + a)\} + \beta\sigma^2_{1,t-1} \\
&= \omega_1 + \alpha\eta^2_{t-1}(1-\mathbb{1}_{t-1})a + (\alpha\eta^2_{t-1} + \beta)\sigma^2_{1,t-1}.
\end{aligned} \tag{13}$$

Thus, provided $\alpha + \beta < 1$, the process is covariance stationary and we have[4]

$$\mathbb{E}[\sigma^2_{1t}] = \frac{\omega_1 + \alpha\, a/2}{1-\alpha-\beta} = \frac{1}{1-\alpha-\beta}\left[\omega_1 + \frac{\omega_2-\omega_1}{2}\frac{\alpha}{1-\beta}\right].$$

An example of such a NIC is shown in Panel (b) of Figure 1.

The sign-switching GARCH model is not very successful empirically (Fornari and Mele, 1997) and a more popular asymmetric process is the one of Glosten et al. (1993), i.e., the GJR-GARCH, where the NIC has its minimum at zero but with different slopes for negative and positive shocks. To reproduce such a shape with model (9) and (8), we may set $\omega_1 = \omega_2 \equiv \omega$, $\beta_{1,1} = \beta_{2,1} = 0$, and $\beta_{1,2} = \beta_{2,2} \equiv \beta$, i.e., we have the restricted *non-diagonal* MixN-GARCH(1,1) model

$$\begin{pmatrix} \sigma^2_{1t} \\ \sigma^2_{2t} \end{pmatrix} = \begin{pmatrix} \omega \\ \omega \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}\varepsilon^2_{t-1} + \begin{pmatrix} 0 & \beta \\ 0 & \beta \end{pmatrix}\begin{pmatrix} \sigma^2_{1,t-1} \\ \sigma^2_{2,t-1} \end{pmatrix}. \tag{14}$$

We may note that, in contrast to the example above, (14) does not admit a diagonal GARCH(1,1) representation, as we observe when we plug the ARCH(∞) representation of $\sigma^2_{2,t-1} = \omega(1-\beta)^{-1} + \alpha_2(1-\beta L)^{-1}\varepsilon^2_{t-2}$ into the equation for σ^2_{1t}, that is,

$$\begin{aligned}
\sigma^2_{1t} &= \omega + \alpha_1\varepsilon^2_{t-1} + \beta\sigma^2_{2,t-1} \\
&= \omega + \alpha_1\varepsilon^2_{t-1} + \frac{\beta\omega}{1-\beta} + \frac{\beta\alpha_2\varepsilon^2_{t-2}}{1-\beta L},
\end{aligned}$$

which through multiplication with $1-\beta L$ shows that the *diagonal* representation of this process is restricted MixN-GARCH(1,2), namely

$$\begin{pmatrix} \sigma^2_{1t} \\ \sigma^2_{2t} \end{pmatrix} = \begin{pmatrix} \omega \\ \omega \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}\varepsilon^2_{t-1} + \begin{pmatrix} \beta(\alpha_2-\alpha_1) \\ 0 \end{pmatrix}\varepsilon^2_{t-2} + \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix}\begin{pmatrix} \sigma^2_{1,t-1} \\ \sigma^2_{2,t-1} \end{pmatrix}.$$

Using (5), the conditional variance for model (14) works out as

$$\sigma^2_t = \lambda_t\sigma^2_{1t} + (1-\lambda_t)\sigma^2_{2t} = \omega + \alpha_1\varepsilon^2_{t-1} + (1-\lambda_t)(\alpha_2-\alpha_1)\varepsilon^2_{t-1} + \beta\sigma^2_{2,t-1}. \tag{15}$$

With $\alpha_2 > \alpha_1$ and $\gamma_1 > 0$ as above, the NIC is centered at zero but has a larger slope for negative shocks than for positive shocks of the same magnitude, as illustrated in Panel (c) of Figure 1.

[4] The unconditional process variance is

$$\mathbb{E}[\varepsilon^2_t] = \frac{1}{2}\mathbb{E}[\sigma^2_{1t}] + \frac{1}{2}\mathbb{E}[\sigma^2_{2t}] = \mathbb{E}[\sigma^2_{1t}] + \frac{a}{2} = \frac{(\omega_1+\omega_2)/2}{1-\alpha-\beta}.$$

The limiting case $\gamma_1 \to \infty$ such that (12) holds would then correspond to the GJR model, except again that the lagged variance in (15) is $\sigma^2_{2,t-1}$ rather than σ^2_{t-1}.[5] Again, for the limiting case, the unconditional expectation of σ^2_{1t} and hence the NIC can be evaluated exactly. To do so, we define the indicator variable $\mathbb{1}_t$ and $\{\eta_t\}$ as in (13), so that we can write

$$\begin{pmatrix} \sigma^2_{1t} \\ \sigma^2_{2t} \end{pmatrix} = \begin{pmatrix} \omega \\ \omega \end{pmatrix} + \begin{pmatrix} \alpha_1 \eta^2_{t-1} \mathbb{1}_{t-1} & \alpha_1 \eta^2_{t-1}(1 - \mathbb{1}_{t-1}) + \beta \\ \alpha_2 \eta^2_{t-1} \mathbb{1}_{t-1} & \alpha_2 \eta^2_{t-1}(1 - \mathbb{1}_{t-1}) + \beta \end{pmatrix} \begin{pmatrix} \sigma^2_{1,t-1} \\ \sigma^2_{2,t-1} \end{pmatrix}.$$

Thus, the process is covariance stationary if the maximal eigenvalue of the matrix

$$\boldsymbol{C} = \begin{pmatrix} \alpha_1/2 & \alpha_1/2 + \beta \\ \alpha_2/2 & \alpha_2/2 + \beta \end{pmatrix}$$

is below unity. It follows from the results of Haas et al. (2004b) that this is equivalent to the condition $\beta < 1$ and

$$\det(\boldsymbol{I}_2 - \boldsymbol{C}) = 1 - \beta - \frac{\alpha_1 + \alpha_2}{2} - \beta \frac{\alpha_2 - \alpha_1}{2} = 1 - \beta - \overline{\alpha} - \beta \frac{\Delta}{2} > 0, \tag{16}$$

where $\overline{\alpha} = (\alpha_1 + \alpha_2)/2$, and $\Delta = \alpha_2 - \alpha_1$.[6] This condition is not identical (although very similar) to that for the corresponding GJR model, i.e., $\beta + (\alpha_1 + \alpha_2)/2 < 1$ (cf. Ling and McAleer, 2002). If condition (16) holds, the unconditional expectation of the component-specific variances is

$$\mathbb{E} \begin{bmatrix} \sigma^2_{1t} \\ \sigma^2_{2t} \end{bmatrix} = (\boldsymbol{I}_2 - \boldsymbol{C})^{-1} \begin{pmatrix} \omega \\ \omega \end{pmatrix} = \frac{1}{1 - \beta - \overline{\alpha} - \beta\Delta/2} \begin{pmatrix} \omega(1 - \Delta/2) \\ \omega(1 + \Delta/2) \end{pmatrix},$$

which can be used to construct the NIC.

In the unrestricted diagonal MixN-GARCH model, where $\beta_{1,2} = \beta_{2,1} = 0$ and $\beta_{1,1} \equiv \beta_1$ and $\beta_{2,2} \equiv \beta_2$ in (14), the conditional variance becomes

$$\begin{aligned} \sigma^2_t &= \sigma^2_{1t} + (1 - \lambda_t)(\sigma^2_{2t} - \sigma^2_{1t}) && (17) \\ &= \omega_1 + \alpha_1 \varepsilon^2_{t-1} + \beta_1 \sigma^2_{1,t-1} + \frac{\omega_2 - \omega_1 + (\alpha_2 - \alpha_1)\varepsilon^2_{t-1} + \beta_2 \sigma^2_{2,t-1} - \beta_1 \sigma^2_{1,t-1}}{1 + \exp\{\gamma_0 + \gamma_1 \varepsilon_{t-1}\}}. && (18) \end{aligned}$$

The simplest possible specification of the form (17) appears when *both* conditional regime-specific variances are constant, i.e., in (19), $\alpha = \beta = 0$, so that $\sigma^2_{jt} = \omega_j$, $j = 1, 2$.[7] With $\omega_1 < \omega_2$, this leads to a NIC which decreases monotonically in a logistic fashion, as illustrated in Panel (d) of Figure 1. The logistic shape of the NIC is not very reasonable, and a more plausible specification with (potentially) monotonically decreasing NIC is obtained when only one of the component variances is constant, which is termed *partial* MixN-GARCH in Haas et al. (2004b), i.e.,

$$\sigma^2_{1t} = \omega_1, \quad \sigma^2_{2t} = \omega_2 + \alpha \varepsilon^2_{t-1} + \beta \sigma^2_{2,t-1}. \tag{19}$$

[5] We decided to have $\sigma^2_{2,t-1}$ rather than $\sigma^2_{1,t-1}$ appear in the component-specific GARCH recursions in (14) due to our assumption that $\alpha_2 > \alpha_1$. It may happen that $\alpha_1 = 0$, i.e., positive shocks have no impact on the conditional volatility. Then, with the roles of $\sigma^2_{1,t-1}$ and $\sigma^2_{2,t-1}$ interchanged in (14), i.e., $\beta_{1,1} = \beta_{2,1} \equiv \beta$ and $\beta_{1,2} = \beta_{2,2} = 0$, σ^2_{1t} would rapidly converge to a constant and σ^2_{2t} would reduce to an ARCH(1) process.

[6] Note that (16) can be rewritten as $\alpha_2 + \beta < 1 + (1 - \beta)\Delta/2$, which shows that the GARCH parameters in the second component need not satisfy the condition $\alpha_2 + \beta < 1$. Conditions for stationarity for the general model are not known. The results of Bauwens et al. (2006) cannot be applied since they assume $\lambda_t \to 1$ as $\varepsilon^2_{t-1} \to \infty$. In such situations, simulation methods as proposed in Gallant et al. (1993) could, in principle, be used to investigate the stationarity of a given model. As far as the properties of the maximum likelihood estimator are concerned, simulations in Cheng et al. (2009) for mixture GARCH models with time-varying weights and typical sample sizes in finance suggest consistency with asymptotic variances being well approximated by the diagonal elements of the inverse of the observed information matrix.

[7] The model is a standard Gaussian mixture with time–varying weight, a special case of the LMARX process of Wong and Li (2001). As these authors show, such models can conveniently be estimated via a small extension of the EM algorithm for standard iid mixture models.

The conditional variance then becomes

$$\sigma_t^2 = \omega_1 + (1-\lambda_t)(\omega_2 - \omega_1 + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{2,t-1}^2) = \omega_1 + \frac{\omega_2 - \omega_1 + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{2,t-1}^2}{1+\exp\{\gamma_0 + \gamma_1\varepsilon_{t-1}\}}. \tag{20}$$

As ε_{t-1} increases, (20) will eventually converge to ω_1, but the convergence may or may not be monotonic. An example for monotonic convergence (i.e., a monotonically decreasing NIC) is provided in Panel (e) of Figure 1. The *unrestricted* diagonal specification with conditional variance as in (17) can capture more complex behaviors of conditional volatility. As an example, consider a pattern emphasized by Fornari and Mele (1997), namely that "high negative shocks increase future volatility more than high positive ones while—at the same time—small positive shocks too often produce a stronger impact on future volatility than negative shocks of the same size", as illustrated in Panel (f) of Figure 1.[8] This occurs when $\alpha_2 > \alpha_1$ as in the GJR-type model but the NIC assumes its minimum at a negative value.

3.2 Time–varying mixture GARCH with likelihood driven mixing weights

The second model takes on lagged likelihood values as the driver of the current mixing weights. In other words, the time conditional process of the mixing weights is driven by the explanatory power of the (mixture) component models based on their historic performance. We consider this a natural link between yesterday's return and today's volatility as the component model that best explains past returns is rewarded with a higher weight, while the other components proportionally receive lower weights (the vector of mixing weights must sum to one). Different domains of expertise (a term referring to expert systems in the field of cognitive systems in computer science) are thus defined, in a non-linear fashion via the component-wise likelihood functions, by higher and lower mixing weights, which lead to a (possibly) asymmetric news impact curve (NIC) regarding the overall variance of the model. This partitioning is further emphasized by different mixture component means, μ_i. The model structure for the mixing weights is given by

$$\lambda_{jt} = \frac{W_{jt}}{\sum_i W_{it}}, \quad W_{jt} = \nu_j + \sum_{m=1}^{u} \gamma_m \frac{\ell_{j,t-m}}{\sum_i \ell_{i,t-m}}, \tag{21}$$

where $\ell_{jt} = \phi\left(r_t; \mu_{jt}, \sigma_{jt}^2\right)$, $\nu_j > 0$, $j = 1, \ldots, k$, and $\gamma_m \geq 0$, $m = 1, \ldots, u$. As in (4), a standard GARCH structure is considered for the k mixture components. Similar to (6), additional terms (lagged terms of λ_{jt} or of exogenous variables) could be entertained to augment (21).[9] We focus, however, on a sparse parametrization and stick to $u = 1$, i.e., $W_{jt} = \nu_j + \gamma\ell_{j,t-1}/(\sum_i \ell_{i,t-1})$. The limiting case for the sparse model, where only $\ell_{p,t-1}$ is different from zero, takes the form

$$\lambda_{jt} = \frac{\nu_j + \gamma\mathbb{1}_{j=p}}{1+\gamma\mathbb{1}_{j=p}}, \tag{22}$$

such that the deviation from $\boldsymbol{\lambda} = \boldsymbol{\nu}$ is bounded above by (22) as a function of γ. A leverage-type effect can evolve as a special form of an asymmetric NIC, e.g., compare Asai and McAleer (2011), if the mixture GARCH components (increasingly ordered by their component means) form a decreasing series concerning the amplitudes of their volatility dynamics. To be precise, by construction of the model, this leverage-type effect is limited to the center of the data, where the NIC can be characterized by the different domains of expertise (or regimes of volatility); whereas the GARCH component of highest volatility will dominate the outer area by the scale of γ in (22). Hence, one may want to call the modeled effect a *partial leverage effect*, although our empirical testing confirms that the effect (if present) usually affects more than 90% of the observed data. Figure 2 shows exemplarily that the effect is indeed found in empirical returns

[8] In Fornari and Mele (1997) the *volatility-switching* GARCH model is designed to reproduce this effect.

[9] Spillover effects, as an integral part of the literature on multivariate (GARCH) models, e.g., see McAleer and da Veiga (2008), may likewise be implemented as in (21) by a likelihood driven linkage.

data, while Section 4 confirms the usefulness of this approach in an exhaustive out-of-sample forecast study. The leverage nature of the effect arises from the change in the mixing weights, such that from a stylized point of view, the high volatility component dominates for (lagged) returns between -3 and 0, while the contrary holds between 0 and 3. Similar patterns are a robust finding in almost all estimates, if $\gamma > 0$.

3.2.1 Estimation using an embedded EM Algorithm

We distinguish between two model variants, namely MixN-GARCH-LIK and MixN-GARCH-LIKW. In the first model, MixN-GARCH-LIK, we restrict vector $\boldsymbol{\nu}$ to be estimated separately from the other parameters, and enforce that $\boldsymbol{\nu}$ maximizes the log likelihood function of the MixN-GARCH model without time-varying mixing weight, i.e.,

$$\boldsymbol{\nu} = \arg\max_{\boldsymbol{\lambda}} \sum_{t=1}^{T} \log \left(\sum_{j=1}^{k} \lambda_j \ell_{jt} \right).$$

In doing so, the accessible parameter space is being shrunk in the sense that the MixN-GARCH model with constant mixing weights becomes the linchpin of the new model, i.e., that when maximizing the likelihood, the new model always nests the *optimal* one with constant weights for $\gamma = 0$. This restriction basically avoids the interaction of component and mixing parameters, except for the *leverage related* γ, and dramatically improves the out-of-sample quality as shown in Table 7 in comparison with the second variant, MixN-GARCH-LIKW, for which all parameters are estimated jointly.

Practically, we estimate $\boldsymbol{\nu}$ ceteris paribus using a *reduced EM* (REM) algorithm derived from the standard EM for mixtures of normals, so that $\boldsymbol{\nu}$ can be used in-place in a nested optimization, eluding a two-step procedure. Let $\boldsymbol{\ell}$ be the $k \times T$ matrix of the component-wise likelihood values (given the current estimate of the component models from the outer estimation of the GARCH parameters and γ). The REM algorithm cuts off the estimation of the component-wise density parameters by leaving them constant and estimates the mixing weights only, i.e., it iterates over

$$\nu_{j,n+1} = \frac{1}{T} \sum_{t=1}^{T} \frac{\nu_{j,n} \ell_{jt}}{\sum_i \nu_{i,n} \ell_{it}},$$

where $\nu_{j,0} = 1/k$ is the initial starting value, and stops if $|\nu_{j,n+1} - \nu_{j,n}| \leq \epsilon$ with, e.g., $\epsilon = 10^{-4}$. Most notably, REM exhibits a linear rate of convergence and has been observed to slow down the nested estimation for $k \geq 3$. It nevertheless remained feasible in all our empirical testing. For preventing the degeneracy of mixture components in the outer estimation, we employ the extended augmented maximum likelihood estimator (EALE) introduced in Broda et al. (2013).

3.3 Asymmetric Mixed Normal GARCH

In order to capture the leverage effect, Alexander and Lazar (2009) propose two asymmetric extensions of the MixN-GARCH model defined by (1) and (2). As we will consider these in our empirical applications below, we introduce them here. The first of these extensions, MixN-GARCH-ASYM, uses the asymmetric GARCH specification of Engle (1990) i.e., the GARCH process driving the variance of mixture component j is given by

$$\sigma_{jt}^2 = \omega_j + \alpha_j (\varepsilon_{t-1} - \theta_j)^2 + \beta_j \sigma_{j,t-1}^2, \quad j = 1, \ldots, k, \tag{23}$$

where the θ_j's are the parameters monitoring the component-specific leverage effect. In particular, if $\theta_j > 0$, a negative shock will increase the next period's σ_{jt}^2 more than a positive shock; a multivariate version of MixN-GARCH-ASYM has been investigated in Haas et al. (2009). The second variant, MixN-GARCH-GJR, employs the model of Glosten et al. (1993), widely known as GJR–GARCH, and specifies the variance process of component j as

$$\sigma_{jt}^2 = \omega_j + \alpha_j \varepsilon_{t-1}^2 + \theta_j d_{t-1}^- \varepsilon_{t-1}^2 + \beta_j \sigma_{j,t-1}^2, \quad j = 1, \ldots, k,$$

where $d_{t-1}^- = 1$ if $\varepsilon_{t-1} < 0$ and $d_{t-1}^- = 0$ otherwise. As in (23), a positive θ_j implies that σ_{jt}^2 reacts more intensely to negative shocks than to positive shocks.

4 Empirical Results

The empirical analysis is based on the major equity indices DAX30, S&P500, DJIA30, NIKKEI225 and NASDAQ COMPOSITE (10 years of data, dating back from July 7th, 2009) as well as the exchange rates JPY/EUR and USD/EUR (5 years of data, dating back from July 7th, 2009). All results (in-sample and out-of-sample) are based on daily percentage log returns, $\varepsilon_t = 100 (\log P_t - \log P_{t-1})$, where P_t is the daily closing level of the index at time t.

As discussed, in this paper, we propose the two new models MixN-GARCH-LIK and MixN-GARCH-LOG, and their modeling and forecasting properties are described in this section. As a brief summary, of the two models it is MixN-GARCH-LIK that outperforms all its competitors by quite a huge margin. In fact, in the many out-of-sample forecasting exercises we discuss below it is this model, MixN-GARCH-LIK, that, for most summary statistics, archives the best scores independent of the datasets considered. For simplicity, we only entertain one and two component models in this paper, but results are also heavily in favor for MixN-GARCH-LIK when comparing three component models.[10]

4.1 In-Sample Fit

For assessing in-sample properties, we fit all models under study to the entire data range, i.e., in Table 1 and 2 we show the likelihood values and BIC measures of all models and data sets. We focus on the BIC because the literature on mixture models provides some theoretical and empirical justification for its appropriateness and good performance, in particular for selecting the number of mixture components (see, e.g., Keribin, 2000; Francq et al., 2001; and Frühwirth-Schnatter, 2010, Ch. 4).

As expected, the pure likelihood values favor the two component models and center around the MixN-GARCH-ASYM and MixN-GARCH-GJR (both models with 11 free parameters). What is (perhaps) surprising is the fact that BIC, which favors less densely parameterized models, also has an overall tendency towards the two component models. In fact, for all data sets, the BIC signals superiority of the two–component models and, of those, the MixN-GARCH-GJR model wins in three out of the seven cases, even though this model has the highest parametrization, with 11 free parameters. However, the BIC of the MixN-GARCH-GJR model is not far from the two newly proposed ones, MixN-GARCH-LIK and MixN-GARCH-LOG, and as mentioned before, it is these models that shine above all in the more recognized out-of-sample forecasting comparison.

4.2 Forecasting Performance

More flexible models (e.g., all types of two component models) should be expected to provide an excellent in-sample fit to virtually any return series compared with more traditional (one component) GARCH-type models including the ones that can model several asymmetries, but the concern remains as to whether the additional parametrization and the nontrivial computational aspects of the feedback between different components warrant its use. To judge this, we compare the empirical performance of the one-step-ahead predictive cdfs across models using tests for

[10]The results are available from the authors on request.

uniformity (see below) as well as a variety of tests concentrating on the left tail of the return distribution as in Broda et al. (2013).

For all models considered, we re-estimate the model parameters every 20 trading days (about once a month), so that each estimation contains 2% of new data. Our analysis is based on the realized predictive cdf values obtained from evaluating the one-step-ahead cdf forecasts at the realized returns. If the model is correct, it is well-known that these are independently and uniformly distributed over the unit interval (Rosenblatt, 1952).

Let $\hat{p}_t = \hat{F}_{t-1}(\varepsilon_t; \hat{\boldsymbol{\theta}}_{t-h})$, $t = 1, \ldots, N$, be the sequence of realized predictive cdf values, noting that, for each t, the parameter vector is estimated using information (in this case, just the past returns) up to and including time $t-h$, where h is a value in $\{1, 2, \ldots, 20\}$, but the entire return series up to time $t-1$ is used in the model filter. Finally, this predictive cdf is evaluated at the actual return at time t. Denote the collection of these N values as vector $\hat{\boldsymbol{p}}$. Further let $\hat{\boldsymbol{p}}^{[s]}$ denote the sorted vector, $\hat{p}_1^{[s]} \leq \hat{p}_2^{[s]} \leq \cdots \leq \hat{p}_N^{[s]}$. The Anderson-Darling (AD) and Cramér-von Mises (CM) test statistics are given respectively by

$$\mathrm{AD} = -N - \sum_{i=1}^{N} \frac{2i-1}{N} \left(\log(\hat{p}_i^{[s]}) + \log(1 - \hat{p}_{N-i+1}^{[s]}) \right)$$

and

$$\mathrm{CM} = \frac{1}{12N} + \sum_{i=1}^{N} \left(\frac{2i-1}{2N} - \hat{p}_i^{[s]} \right)^2 .$$

In addition, we provide test statistics for the Kolmogorov-Smirnov (KS) test for uniformity, as well as the Jarque-Bera (JB) and Shapiro-Wilk (SW) tests for normality. It is important to note that we are testing the prediction quality over the whole support of the distribution, and not just the left tail (as we do below, for directly testing the quality of value at risk predictions). Table 3 shows the results. The statistics AD, CM and KS reveal the astonishing performance of model MixN-GARCH-LIK in comparison to its competitors: in four out of seven cases for AD and in three out of seven cases for both CM and KS, it is MixN-GARCH-LIK that scores highest. For JB and SW there is also a clear tendency towards the two component models but no obvious pattern towards a particular type arises.

We also consider VaR measures dedicated to the left tail, as these are of even greater interest from a risk management perspective. Table 4 shows the empirical coverage probabilities (as percentages) for the 1% and 5% VaR levels. The results in Table 4 confirm the superiority of the proposed models in four out of seven cases at the 1% level and three cases at the 5% level.

For further investigations of the VaR prediction quality, we adopt a simple quality measure based on the coverage error over the VaR levels up to $100\lambda\%$, see Kuester et al. (2006). The measure calculates the deviation between predictive cdf and uniform cdf and, thus, captures the excess of percentage violations over the VaR levels, where the deviation is defined as $100(F_U - \hat{F}_e)$ with F_U being the cdf of the standard uniform random variable and $\hat{F}_e$ referring to the empirical cdf formed from $\hat{\boldsymbol{p}}$. Building upon this metric we report the integrated root mean squared error (IRMSE) over the left tail up to the maximal VaR level of interest. The IRMSE employed herein is closely related to the CM statistic but with the sum truncated at $h = \lceil \lambda N \rceil$, i.e.,

$$\mathrm{IRMSE} = \sqrt{\frac{1}{h} \sum_{i=1}^{h} \left(100 \frac{2i-1}{2N} - 100 \hat{p}_i^{[s]} \right)^2} .$$

The results on the IRMSE in Table 5 are also in line with our general observation that MixN-GARCH-LIK is the overall winning model.

Finally as in Broda et al. (2013), we also investigate the hit sequence of realized predictive VaR violations,

$$v_t = \mathbb{1}_{\varepsilon_t \leq \hat{q}_t}, \quad \hat{q}_t = \widehat{\mathrm{VaR}}_{t|t-1}(\lambda),$$

where $\mathbb{1}$ is the indicator function. Under the null of correct conditional coverage, the v_i are iid Bernoulli(λ). From this sequence, the test statistic $\mathrm{LR}_{\mathrm{CC}} = \mathrm{LR}_{\mathrm{UC}} + \mathrm{LR}_{\mathrm{IND}}$ is computed, as proposed in Christoffersen (1998), where $\mathrm{LR}_{\mathrm{UC}}$ and $\mathrm{LR}_{\mathrm{IND}}$ test for unconditional coverage and independence, respectively. As can be seen from Table 6 for the 1% VaR level, as well as for the 5% VaR level, for all tests a clear tendency arises toward MixN-GARCH-LIK. With only one exception MixN-GARCH-LIK is the best performing model for every test.

5 Conclusions and Further Extensions

In this paper, we relax the assumption of constant weights in the class of mixed normal GARCH processes and introduce two different flexible time–varying weight model structures. Current mixing weights (and hence the implied overall volatility) are either directly related to past innovations by logistic response functions or indirectly via their lagged component-specific likelihood contributions. In particular, the second model type allows non-linear feedback between its likelihood components, and so induces news impact curves with (partial) leverage effects. As demonstrated, this latter model delivers clear-cut superior out-of-sample performance compared to all entertained models; and this, over a variety of data sets. Important open issues to be addressed in future research include establishing the stationarity conditions for the model and the asymptotic properties of the (augmented) maximum likelihood estimator.

The model classes are quite rich, and future applications should entertain other choices of the parameters and form structures. As mentioned above, Engle and Ng (1993) show that the older the news, the smaller the impact on current and future volatility. Also for the leverage effect, it is well known from Bouchaud et al. (2001) that its decay time differs across assets, with stocks requiring about 10 days, and indices about 50 days. These authors also show that the serial correlation function describing the magnitude of the leverage effect in terms of lags can be fit with a (single) exponential, which can be directly related to the dynamics between current mixing weights and past model innovations.

In addition, more flexible and more asymmetric model structures might be useful in order to further account for the "down–market effect" or "panic effect", i.e., a "one–sided" leverage effect related to falling stock prices. In fact, according to Figlewski and Wang (2000), a rise in the stock price does not affect volatility at all. They find the leverage effect is just a "down-market effect" not being existent for positive news surprises. This can easily be incorporated in our models by extending the constant weight assumption just for negative innovations, and/or using non-parametric response functions. Moreover, in addition to, or instead of, relating current mixing weights to the past innovations and likelihood contributions, it might be advantageous to consider use of the conditional variance, skewness or kurtosis. Further improvements to forecasting performance could also be gained by use of weighted likelihood; see Paolella and Steude (2008).

Finally, extensions into a multivariate framework are possible. For example, a straightforward generalization of MixN-GARCH-LIK is derived, e.g., by using the multivariate mixture GARCH model in Haas et al. (2009), given that the process of the mixture weights in (21) is entirely likelihood driven and, hence, generally applicable to univariate as well as multivariate models. Alternatively, it appears possible to augment the EM algorithm approach used for the multivariate mixture-based GARCH model in Paolella and Polak (2013) to the model structure used herein, thus rendering estimation in high dimensions feasible. Another approach which is also feasible in high dimensions is the use of Independent Component Analysis methods. Given the tractability of the moment generating and characteristic function of the conditional mixed normal forecast distribution used in this paper, the methodology in Broda and Paolella (2009) and Broda et al. (2013) is directly applicable. These ideas are currently being pursued.

References

Alexander, C. (2008). *Practical Financial Econometrics.* John Wiley & Sons, Chichester.

Alexander, C. and Lazar, E. (2006). Normal Mixture GARCH(1,1): Applications to Exchange Rate Modelling. *Journal of Applied Econometrics*, 21:307–336.

Alexander, C. and Lazar, E. (2009). Modelling Regime–specific Stock Price Volatility. *Oxford Bulletin of Economics and Statistics*, 71:761–797.

Aparicio, F. M. and Estrada, J. (2001). Empirical Distributions of Stock Returns: European Securities Markets, 1990–95. *European Journal of Finance*, 7:1–21.

Asai, M. and McAleer, M. (2011). Alternative Asymmetric Stochastic Volatility Models. *Econometric Reviews*, 30(5):548–564.

Bauwens, L., Preminger, A., and Rombouts, J. (2006). Regime Switching GARCH Models. CORE Discussion Paper 2006/11, Université Catholique de Louvain.

Bekaert, G. and Gray, S. F. (1998). Target Zones and Exchange Rates: An Empirical Investigation. *Journal of International Economics*, 45:1–35.

Black, F. (1976). Studies in Stock Price Volatility Changes. In *American Statistical Association, Proceedings of the Business and Economic Statistics Section*, pages 177–181.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Bouchaud, J.-P., Matacz, A., and Potters, M. (2001). Leverage Effect in Financial Markets: The Retarded Volatility Model. *Physical Review Letters*, 87(22):1–4.

Broda, S. A., Haas, M., Krause, J., Paolella, M. S., and Steude, S. C. (2013). Stable mixture GARCH models. *Journal of Econometrics*, 172(2):292–306.

Broda, S. A. and Paolella, M. S. (2009). CHICAGO: A Fast and Accurate Method for Portfolio Risk Calculation. *Journal of Financial Econometrics*, 7(4):412–436.

Cheng, X., Yu, P. L. H., and Li, W. K. (2009). On a Dynamic Mixture GARCH Model. *Journal of Forecasting*, 28:247–265.

Christie, A. (1982). The Stochastic Behavior of Common Stock Variance: Value, Leverage, and Interest Rate Effects. *Journal of Financial Economics*, 10:407–432.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39(4):841–862.

Engle, R. F. (1990). Stock Volatility and the Crash of '87: Discussion. *Review of Financial Studies*, 3:103–106.

Engle, R. F. and Ng, V. K. (1993). Measuring and Testing the Impact of News on Volatility. *The Journal of Finance*, 48:1749–1778.

Fama, E. F. (1965). The Behavior of Stock Market Prices. *Journal of Business*, 38:34–105.

Figlewski, S. and Wang, X. (2000). Is the "Leverage Effect" a Leverage Effect? Working paper series 00-37, New York University.

Fornari, F. and Mele, A. (1997). Sign- and Volatility-Switching ARCH Models: Theory and Applications to International Stock Markets. *Journal of Applied Econometrics*, 12:49–65.

Francq, C., Roussignol, M., and Zakoïan, J.-M. (2001). Conditional Heteroskedasticity Driven by Hidden Markov Chains. *Journal of Time Series Analysis*, 22(2):197–220.

Frühwirth-Schnatter, S. (2010). *Finite Mixture and Markov Switching Models*. Springer.

Gallant, A. R., Rossi, P. E., and Tauchen, G. (1993). Nonlinear Dynamic Structures. *Econometrica*, 61(4):871–907.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5):1779–1801.

Haas, M., Mittnik, S., and Mizrach, B. (2006a). Assessing Central Bank Credibility During the EMS Crises: Comparing Option and Spot Market-Based Forecasts. *Journal of Financial Stability*, 2:28–54.

Haas, M., Mittnik, S., and Paolella, M. S. (2004a). A New Approach to Markov Switching GARCH Models. *Journal of Financial Econometrics*, 2(4):493–530.

Haas, M., Mittnik, S., and Paolella, M. S. (2004b). Mixed Normal Conditional Heteroskedasticity. *Journal of Financial Econometrics*, 2(2):211–250.

Haas, M., Mittnik, S., and Paolella, M. S. (2009). Asymmetric Multivariate Normal Mixture GARCH. *Computational Statistics and Data Analysis*, 53:2129–2154.

Haas, M., Mittnik, S., Paolella, M. S., and Steude, S. C. (2006b). Analyzing and Exploiting Asymmetries in the News Impact Curve. *National Centre of Competence in Research Financial Valuation and Risk Management, Working Paper Series*.

Haas, M. and Paolella, M. S. (2012). Mixture and Regime-switching GARCH Models. In Bauwens, L., Hafner, C. M., and Laurent, S., editors, *Handbook of volatility models and their applications*, number 3.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57:357–384.

Hasanhodzic, J. and Lo, A. W. (2011). Black's Leverage Effect is Not Due to Leverage. *SSRN eLibrary*.

Hens, T. and Steude, S. C. (2009). The leverage effect without leverage. *Finance Research Letters*, 6(2):83–94.

Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhyā: The Indian Journal of Statistics, Series A*, 62(1):49–66.

Kim, T.-H. and White, H. (2004). On More Robust Estimation of Skewness and Kurtosis. *Finance Research Letters*, 1:56–73.

Klaster, M. A. and Knot, K. H. W. (2002). Toward an Econometric Target Zone Model with Endogenous Devaluation Risk for a Small Open Economy. *Economic Modelling*, 19:509–529.

Kon, S. J. (1984). Models of Stock Returns: A Comparison. *The Journal of Finance*, 39:147–165.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value–at–Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics*, 4:53–89.

Ling, S. and McAleer, M. (2002). Stationarity and the Existence of Moments of a Family of GARCH Processes. *Journal of Econometrics*, 106:109–117.

McAleer, M. and da Veiga, B. (2008). Forecasting value-at-risk with a parsimonious portfolio spillover GARCH (PS-GARCH) model. *Journal of Forecasting*, 27(1):1–19.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley.

Neely, C. J. (1999). Target Zones and Conditional Volatility: The Role of Realignments. *Journal of Empirical Finance*, 6:177–192.

Neftci, S. N. (2000). Value at Risk Calculations, Extreme Events, and Tail Estimation. *Journal of Derivatives*, 7:23–28.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59:347–37.

Palm, F. C. (1996). GARCH Models of Volatility. In Maddala, G. and Rao, C., editors, *Handbook of Statistics*, volume 14, pages 209–240. Elsevier Science, Amsterdam.

Paolella, M. S. and Polak, P. (2013). COMFORT-CCClass: A Common Market Factor Non-Gaussian Returns Model. submitted.

Paolella, M. S. and Steude, S. C. (2008). Risk prediction: a DWARF-like approach. *The Journal of Risk Model Validation*, 2(1):25–43.

Park, B.-J. (2011). Asymmetric herding as a source of asymmetric return volatility. *Journal of Banking & Finance*, 35(10):2657–2665.

Rodriguez, M. J. and Ruiz, E. (2012). Revisiting Several Popular GARCH Models with Leverage Effect: Differences and Similarities. *Journal of Financial Econometrics*, 10:637–668.

Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics*, 23:470–472.

Tashman, A. and Frey, R. J. (2009). Modeling risk in arbitrage strategies using finite mixtures. *Quantitative Finance*, 9:495–503.

Tucker, A. L. and Pond, L. (1988). The Probability Distribution of Foreign Exchange Price Changes: Tests of Candidate Processes. *Review of Economics and Statistics*, 70:638–647.

Vlaar, P. J. G. and Palm, F. C. (1993). The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroscedasticity, and Jumps. *Journal of Business and Economic Statistics*, 11(3):351–360.

Wong, C. S. and Li, W. K. (2001). On a Logistic Mixture Autoregressive Model. *Biometrika*, 88:833–846.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955.

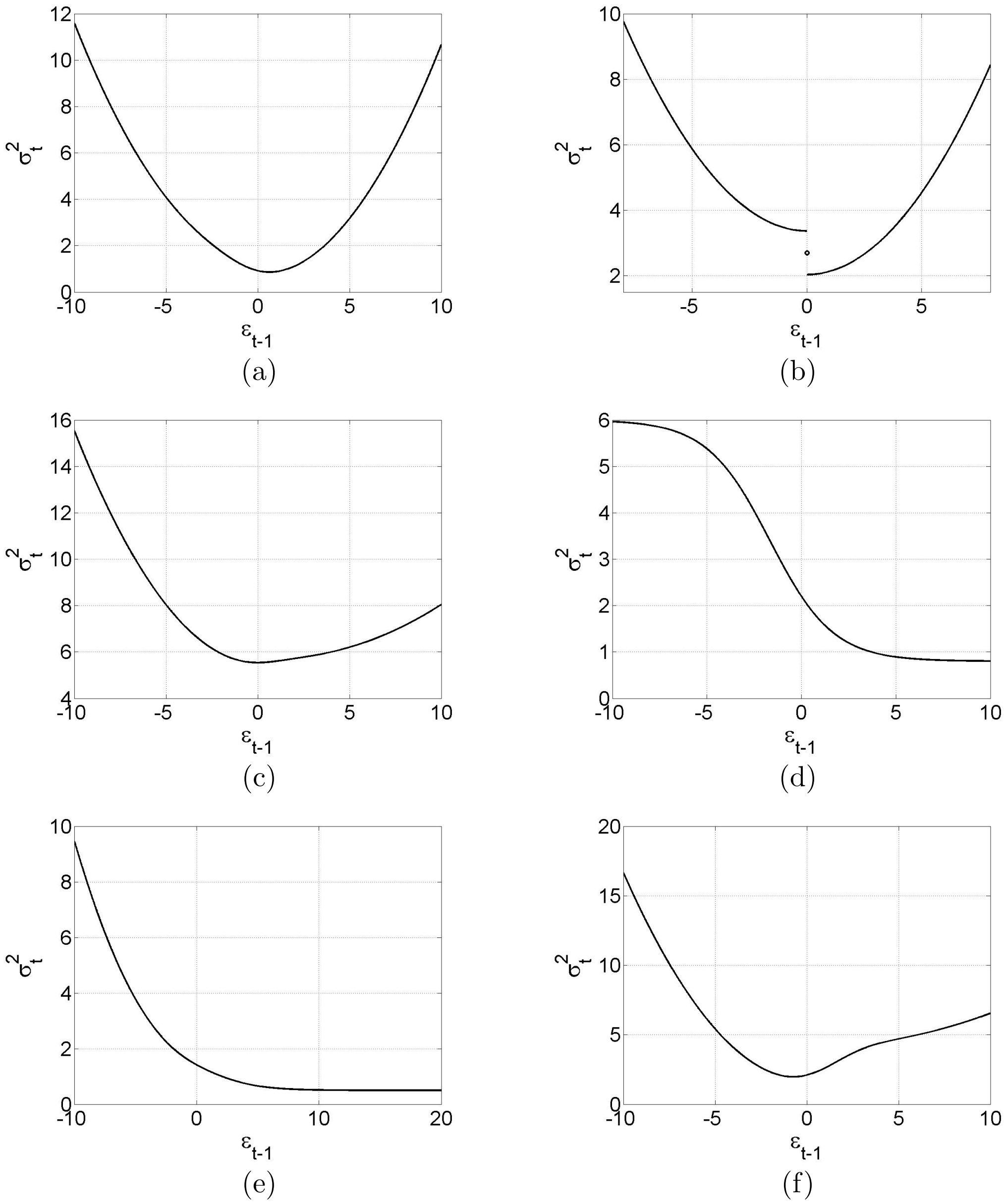

Figure 1: Shown are various possible shapes of the news impact curve (NIC; x-axis ε_{t-1}, y-axis σ_t^2) implied by the MixN-GARCH model (9) with time–varying weights described by (8). Panel (a) shows the NIC for the model with switching intercept (10) with $\gamma_0 = 1$, $\gamma_1 = 1, \omega_1 = 0.01, \omega_2 = 0.15, \alpha = 0.1$, and $\beta = 0.85$. Panel (b) illustrates the limiting case of this model with $\gamma_0 = 0, \gamma_1 = \infty, \omega_1 = 0.05, \omega_2 = 0.25, \alpha = 0.1$, and $\beta = 0.85$. Panel (c) shows the NIC for model (14) and $\gamma_0 = -1, \gamma_1 = 1, \omega = 0.02, \alpha_1 = 0.03, \alpha_2 = 0.1$, and $\beta = 0.9$. Panel (d) shows the NICs for the simple Gaussian mixture (constant variances) with $\gamma_0 = 1, \gamma_1 = 0.6, \omega_1 = 0.8, \omega_2 = 6$, and $\alpha = \beta = 0$. Panel (e) displays the NIC for the partial model (19) with $\gamma_0 = -0.5, \gamma_1 = 0.7, \omega_1 = 0.5, \omega_2 = 0.1, \alpha = 0.08$, and $\beta = 0.9$, and Panel (f) pertains to the general diagonal specification (17) with $\gamma_0 = -2, \gamma_1 = 1, \omega_1 = 0.35, \omega_2 = 0.05, \alpha_1 = 0.03, \alpha_2 = 0.15, \beta_1 = 0.9$, and $\beta_2 = 0.8$.

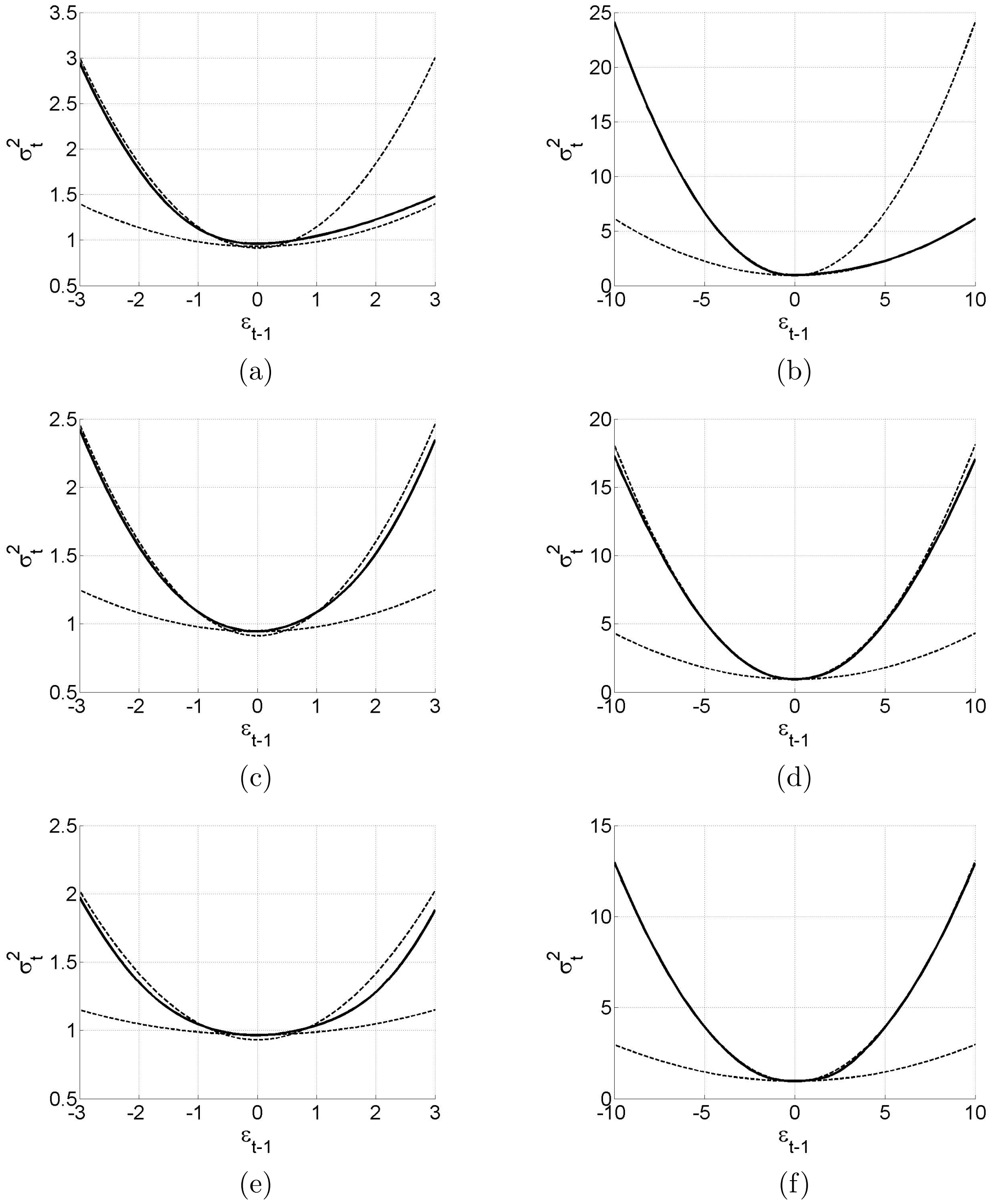

Figure 2: News impact curves (NIC) for selected estimates of MixN-GARCH-LIK and MixN-GARCH-LOG from the out-of-sample forecast exercise in Section 4. Figures (y-axis σ_t^2, x-axis ε_{t-1}) on the left show the magnified center of the figures on the right. Bold lines denote the NIC of the overall mixture, dashed lines denote the component-wise NICs. For MixN-GARCH-LIK, the leverage effect is particularly present in the range $-3 \leq \varepsilon_{t-1} \leq 3$, where for all data sets under study at least 90% of the (percentage log-) returns are located. Panel (a) and (b) are based on MixN-GARCH-LOG, the remaining panels on MixN-GARCH-LIK. Panel (a)–(d) show NICs for the DAX30 returns data as used in Table 3, whereas panel (e) and (f) use NASDAQ COMPOSITE data. The estimated parameter are $\gamma_0 = -0.21, \gamma_1 = -1.05, \mu_1 = -0.33, \mu_2 = 0.21, \omega_1 = 0.07, \omega_2 = 0.002, \alpha_1 = 0.23, \alpha_2 = 0.05, \beta_1 = 0.85, \beta_2 = 0.93$ for panel (a/b), $\gamma = 0.77, \mu_1 = 0.26, \mu_2 = -0.28, \omega_1 = 0.001, \omega_2 = 0.04, \alpha_1 = 0.03, \alpha_2 = 0.17, \beta_1 = 0.94, \beta_2 = 0.87$ for (c/d), and $\gamma = 7.02, \mu_1 = 0.22, \mu_2 = -0.16, \omega_1 = 0.001, \omega_2 = 0.03, \alpha_1 = 0.02, \alpha_2 = 0.12, \beta_1 = 0.97, \beta_2 = 0.9$ for (e/f).

model	free param.	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Normal-GARCH	4	-6089.67	-5254.12	-5143.05	-6334.56	-2372.40	-2220.49	-6387.07
Normal-ASYM-GARCH	5	-6050.43	-5193.92	-5086.38	-6285.86	-2370.69	-2220.22	-6357.60
Normal-GJR-GARCH	5	-6050.10	-5180.52	-5086.28	-6298.24	-2370.20	-2220.49	-6350.38
Normal-EGARCH	5	-6043.56	-5174.52	-5076.13	-6291.27	-2366.82	-2222.25	-6351.47
MixN-GARCH	9	-6036.59	-5190.73	-5076.31	-6277.75	-2332.27	-2212.91	-6355.07
MixN-GARCH-ASYM	11	-6010.95	-5140.66	-5035.04	**-6245.84**	-2328.32	**-2209.64**	-6324.46
MixN-GARCH-GJR	11	**-6003.13**	**-5122.94**	**-5028.65**	-6251.15	-2329.19	-2211.10	**-6309.50**
MixN-GARCH-LIK	10	-6031.16	-5188.23	-5073.16	-6277.30	-2330.67	-2209.86	-6349.75
MixN-GARCH-LOG	11	-6024.64	-5181.35	-5067.54	-6274.52	**-2326.73**	-2210.12	-6335.01

Table 1: In-sample likelihood values for all single- and multi-component mixture GARCH models and all data sets under study. In-sample statistics are based on complete data sets as used in Table 3. For comparison, all models include a location parameter for the density. Entries in boldface denote the best results per data set.

model	free param.	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Normal-GARCH	4	12212.11	10541.01	10318.87	12701.89	4775.77	**4471.95**	12806.90
Normal-ASYM-GARCH	5	12141.82	10428.79	10213.71	12612.68	4780.10	4479.16	12756.15
Normal-GJR-GARCH	5	12141.15	10402.00	10213.52	12637.44	4779.12	4479.69	12741.72
Normal-EGARCH	5	12128.07	10390.00	10193.21	12623.50	4772.36	4483.21	12743.89
MixN-GARCH	9	12138.71	10446.99	10218.16	12621.03	**4726.48**	4487.76	12775.67
MixN-GARCH-ASYM	11	12112.00	10371.41	10160.19	**12581.78**	4741.80	4504.44	12739.02
MixN-GARCH-GJR	11	**12096.37**	**10335.98**	**10147.40**	12592.41	4743.54	4507.37	**12709.10**
MixN-GARCH-LIK	10	12136.05	10450.18	10220.04	12628.33	4731.03	4489.40	12773.23
MixN-GARCH-LOG	11	12131.18	10444.62	10217.00	12630.95	4730.88	4497.66	12751.93

Table 2: BIC values for all single- and multi-component mixture GARCH models and all data sets under study. In-sample statistics are based on complete data sets as used in Table 3 For comparison, all models include a location parameter for the density. Entries in boldface denote the best results per data set.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Anderson-Darling							
Normal-GARCH	4.24***	3.70**	3.37**	3.29**	7.03***	3.37**	2.86**
Normal-ASYM-GARCH	3.94***	3.16**	2.75**	2.44*	6.79***	3.14**	2.73**
Normal-GJR-GARCH	4.01***	2.75**	2.47*	2.84**	6.86***	3.35**	2.70**
Normal-EGARCH	4.66***	3.17**	2.46*	2.70**	6.41***	2.91**	2.77**
MixN-GARCH	1.01	0.91	1.23	1.16	**0.65**	0.62	1.53
MixN-GARCH-ASYM	1.18	1.15	1.12	1.42	0.75	1.16	1.37
MixN-GARCH-GJR	1.12	1.31	1.18	1.20	0.69	1.04	**1.20**
MixN-GARCH-LIK	**0.84**	**0.57**	**0.95**	**1.11**	0.66	**0.61**	1.26
MixN-GARCH-LOG	2.67**	2.16*	2.08*	1.25	0.83	1.07	3.08**
Cramér-von Mises							
Normal-GARCH	0.81***	0.64**	0.61**	0.57**	1.17***	0.69**	0.54**
Normal-ASYM-GARCH	0.73**	0.51**	0.47**	0.37*	1.14***	0.65**	0.54**
Normal-GJR-GARCH	0.72**	0.45*	0.43*	0.47**	1.15***	0.69**	0.51**
Normal-EGARCH	0.72**	0.37*	0.32	0.44*	1.07***	0.61**	0.50**
MixN-GARCH	0.13	0.10	0.18	0.19	**0.07**	0.10	0.18
MixN-GARCH-ASYM	0.14	0.15	0.15	0.20	0.07	0.20	0.21
MixN-GARCH-GJR	0.13	0.19	0.15	0.17	0.07	0.18	0.18
MixN-GARCH-LIK	**0.11**	**0.07**	**0.14**	0.19	0.07	**0.10**	**0.15**
MixN-GARCH-LOG	0.32	0.25	0.30	**0.14**	0.07	0.19	0.38*
Kolmogorov-Smirnov (test statistics are scaled up by factor 100)							
Normal-GARCH	3.93***	3.32**	3.12**	3.01**	6.32***	5.05***	3.41***
Normal-ASYM-GARCH	3.60***	2.79*	2.84*	2.53	6.39***	4.99***	3.32**
Normal-GJR-GARCH	3.70***	2.96**	2.88*	2.90**	6.34***	5.05***	3.26**
Normal-EGARCH	3.22**	2.27	2.54	2.72*	6.04***	4.97***	3.09**
MixN-GARCH	1.77	1.49	2.08	1.89	**1.69**	**2.30**	1.97
MixN-GARCH-ASYM	2.07	1.45	1.92	2.18	1.94	3.07	**1.82**
MixN-GARCH-GJR	1.78	1.78	1.79	2.19	1.97	3.02	1.85
MixN-GARCH-LIK	**1.65**	**1.23**	**1.76**	1.90	1.76	2.37	1.97
MixN-GARCH-LOG	2.32	2.58	2.72*	**1.56**	1.91	2.96	2.43
Jarque-Bera							
Normal-GARCH	243.98***	272.92***	316.41***	163.89***	310.80***	41.75***	89.09***
Normal-ASYM-GARCH	106.97***	208.96***	164.07***	119.16***	338.56***	43.00***	134.64***
Normal-GJR-GARCH	109.15***	319.48***	265.59***	149.15***	312.85***	45.41***	114.31***
Normal-EGARCH	102.35***	320.04***	256.53***	114.18***	274.67***	48.70***	155.01***
MixN-GARCH	38.35***	**38.57*****	35.79***	4.45	**15.57*****	4.54	35.31***
MixN-GARCH-ASYM	**11.14*****	40.72***	**26.33*****	15.16***	51.59***	10.08**	48.45***
MixN-GARCH-GJR	23.04***	64.59***	38.78***	12.22***	30.67***	8.53**	43.99***
MixN-GARCH-LIK	25.57***	40.92***	33.76***	**3.76**	15.94***	**4.48**	**17.99*****
MixN-GARCH-LOG	57.72***	68.96***	72.89***	27.93***	22.68***	21.38***	37.40***
Shapiro-Wilk (test statistic ν is transformed by $1000(1-\nu)$)							
Normal-GARCH	10.24***	11.30***	12.14***	8.80***	29.83***	6.18***	6.12***
Normal-ASYM-GARCH	7.55***	9.72***	8.70***	7.12***	30.37***	6.42***	7.56***
Normal-GJR-GARCH	7.89***	11.55***	10.64***	8.18***	29.67***	6.51***	6.96***
Normal-EGARCH	7.62***	12.28***	10.84***	6.99***	28.42***	6.70***	8.23***
MixN-GARCH	3.52***	**3.19*****	3.36***	1.85***	**4.26*****	1.31	4.03***
MixN-GARCH-ASYM	**2.16*****	3.37***	**2.73*****	2.42***	8.27***	2.36	4.23***
MixN-GARCH-GJR	2.70***	4.40***	3.23***	2.24***	6.38***	2.11	4.17***
MixN-GARCH-LIK	2.93***	3.34***	3.31***	**1.72****	4.33***	**1.29**	**2.96*****
MixN-GARCH-LOG	4.22***	4.36***	4.66***	3.37***	5.62***	3.67***	3.60***

Table 3: Anderson-Darling, Cramér-von Mises, Kolmogorov-Smirnov, Jarque-Bera and Shapiro-Wilk test statistics for all models and data sets under study. Entries in boldface denote the best outcomes. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. All values are based on evaluating the one-step-ahead out-of-sample distribution forecasts at the observed return data, estimated throughout a rolling window of 1,000 trading days, covering 10 years of equity returns and 5 years of FX returns, dating back from July 7, 2009. For $k > 2$, the model parameters are updated every 20 days, while for single-component models, we update in every step as, otherwise, EGARCH would suffer from non-stationary issues preventing a comparison.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
1% VaR							
Normal-GARCH	1.59***	1.77***	1.68***	1.49**	2.48***	1.10	1.18
Normal-ASYM-GARCH	1.71***	1.89***	1.83***	1.88***	2.38***	1.12	1.26
Normal-GJR-GARCH	1.62***	1.76***	1.60***	1.65***	2.40***	1.09	1.29*
Normal-EGARCH	1.81***	1.93***	2.06***	1.86***	2.71***	1.25	1.41**
MixN-GARCH	1.09	1.24	1.21	0.87	1.64**	1.09	**1.00**
MixN-GARCH-ASYM	1.07	1.48**	1.60***	1.29*	1.65**	**1.03**	1.11
MixN-GARCH-GJR	1.15	1.47**	1.32*	1.18	**1.52****	1.05	0.91
MixN-GARCH-LIK	**1.02**	**1.16**	**1.16**	0.78	1.68**	1.09	0.95
MixN-GARCH-LOG	1.31*	1.50**	1.41**	**1.12**	1.87***	1.14	1.03
5% VaR							
Normal-GARCH	5.98**	5.62*	5.45	**5.57**	5.74	4.68	5.94**
Normal-ASYM-GARCH	6.44***	5.57	5.19	5.82**	5.68	4.86	5.79**
Normal-GJR-GARCH	6.34***	5.64*	**5.19**	5.64*	5.66	4.68	5.81**
Normal-EGARCH	6.93***	6.35***	5.61*	5.59*	5.78	5.29	6.05***
MixN-GARCH	5.75**	5.39	5.50	5.77**	5.66	5.00	5.91**
MixN-GARCH-ASYM	6.10***	5.68*	5.45	6.01**	**5.19**	4.72	5.62*
MixN-GARCH-GJR	5.59*	5.91**	5.44	5.80**	5.52	5.05	**5.27**
MixN-GARCH-LIK	**5.37**	**4.84**	5.25	5.72*	5.66	5.03	5.48
MixN-GARCH-LOG	6.30***	5.93**	5.76**	6.08***	5.46	**5.00**	6.64***

Table 4: Predicted VaR coverage percentages (point estimates) at the 1% and 5% level for all models under study. Entries in boldface denote the best (closest to the true value) estimate. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Results are based on the same out-of-sample exercise as in Table 3.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
1% VaR							
Normal-GARCH	0.32	0.38	0.39	0.30	0.52	0.24	0.23
Normal-ASYM-GARCH	0.33	0.43	0.43	0.35	0.52	0.33	0.26
Normal-GJR-GARCH	0.39	0.35	0.41	0.38	0.52	0.25	0.23
Normal-EGARCH	0.45	0.47	0.48	0.35	0.53	0.27	0.29
MixN-GARCH	**0.05**	0.21	0.18	0.12	0.31	0.13	0.08
MixN-GARCH-ASYM	0.10	0.25	0.26	0.21	0.31	**0.10**	0.11
MixN-GARCH-GJR	0.20	**0.14**	0.20	**0.08**	**0.24**	0.18	0.08
MixN-GARCH-LIK	0.05	0.18	**0.16**	0.13	0.31	0.13	**0.06**
MixN-GARCH-LOG	0.21	0.29	0.28	0.19	0.34	0.22	0.12
5% VaR							
Normal-GARCH	0.78	0.80	0.68	0.58	1.17	0.28	0.49
Normal-ASYM-GARCH	0.77	0.83	0.64	0.75	1.19	0.22	0.47
Normal-GJR-GARCH	0.69	0.81	0.69	0.67	1.22	0.26	0.50
Normal-EGARCH	1.02	1.08	0.82	0.70	1.35	0.22	0.68
MixN-GARCH	0.35	0.39	0.44	0.38	0.53	0.19	0.37
MixN-GARCH-ASYM	0.39	0.54	0.53	0.68	0.41	0.18	0.26
MixN-GARCH-GJR	0.29	0.47	0.43	0.53	**0.41**	0.16	**0.10**
MixN-GARCH-LIK	**0.14**	**0.13**	**0.22**	**0.29**	0.53	0.20	0.21
MixN-GARCH-LOG	0.64	0.61	0.61	0.60	0.66	**0.16**	0.53

Table 5: Integrated root mean squared error of the VaR prediction up to the 1% and 5% level for all models under study. Entries in boldface denote the best estimate. Results are based on the same out-of-sample exercise as in Table 3.

	level	model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Unconditional Coverage, LR_{UC}	1%	Normal-GARCH	7.33***	12.51***	10.30***	5.60**	19.81***	0.07	0.88
		Normal-ASYM-GARCH	11.38***	16.15***	14.89***	16.15***	18.02***	0.28	1.71
		Normal-GJR-GARCH	8.27***	12.51***	8.27***	9.26***	18.02***	0.07	2.21
		Normal-EGARCH	13.68***	17.45***	23.05***	14.89***	25.57***	0.63	4.08**
		MixN-GARCH	0.14	1.26	1.26	0.39	4.14**	0.07	**3.1e-4**
		MixN-GARCH-ASYM	0.14	5.60**	8.27***	2.21	5.16**	**1.2e-4**	0.32
		MixN-GARCH-GJR	0.56	4.81**	2.21	0.88	**3.23***	0.07	0.17
		MixN-GARCH-LIK	**0.03**	**0.56**	**0.56**	1.56	5.16**	0.07	0.05
		MixN-GARCH-LOG	2.21	5.60**	4.08**	**0.32**	7.45***	0.28	0.03
	5%	Normal-GARCH	4.97**	2.13	1.05	**1.65**	1.48	0.29	4.60**
		Normal-ASYM-GARCH	10.47***	1.65	0.25	3.57*	1.20	0.08	3.25*
		Normal-GJR-GARCH	8.92***	2.13	**0.17**	2.13	1.20	0.29	3.57*
		Normal-EGARCH	18.50***	9.42***	1.88	1.88	1.48	0.23	5.75**
		MixN-GARCH	2.95*	0.88	1.44	2.95*	1.20	**6.5e-4**	4.24**
		MixN-GARCH-ASYM	6.17**	2.39	1.05	5.36**	**0.12**	0.17	2.13
		MixN-GARCH-GJR	1.88	4.24**	1.05	3.25*	0.72	0.01	**0.34**
		MixN-GARCH-LIK	**0.72**	**0.16**	0.34	2.66	1.20	0.01	1.23
		MixN-GARCH-LOG	8.43***	4.60**	2.95*	6.17**	0.53	**6.5e-4**	13.31***
Independence, LR_{IND}	1%	Normal-GARCH	0.21	0.08	1.54	**0.28**	3.98**	0.33	0.77
		Normal-ASYM-GARCH	0.10	**0.04**	1.84	1.91	4.29**	0.37	0.87
		Normal-GJR-GARCH	0.17	0.08	1.41	1.48	4.29**	0.33	0.92
		Normal-EGARCH	**0.06**	1.99	2.33	1.84	1.02	0.42	1.09
		MixN-GARCH	1.06	0.82	0.82	0.43	0.92	0.33	0.54
		MixN-GARCH-ASYM	0.63	1.21	**0.17**	0.92	0.79	**0.28**	0.67
		MixN-GARCH-GJR	0.72	1.15	0.92	0.77	0.65	0.33	**0.46**
		MixN-GARCH-LIK	1.17	0.72	0.72	0.32	0.79	0.33	0.50
		MixN-GARCH-LOG	0.56	1.21	1.09	0.67	**0.58**	0.37	0.59
	5%	Normal-GARCH	4.40**	0.13	0.33	**0.11**	**0.79**	0.55	0.31
		Normal-ASYM-GARCH	4.50**	2.88*	3.39*	3.81*	1.90	0.40	1.20
		Normal-GJR-GARCH	**2.62**	0.18	0.70	1.80	1.90	0.55	0.28
		Normal-EGARCH	3.46*	2.75*	**0.12**	3.00*	1.73	2.88*	1.84
		MixN-GARCH	3.53*	0.13	1.30	0.17	1.90	0.29	0.22
		MixN-GARCH-ASYM	4.87**	5.16**	2.51	4.56**	1.70	0.47	1.80
		MixN-GARCH-GJR	7.39***	6.18**	2.51	3.67*	1.13	**0.24**	0.34
		MixN-GARCH-LIK	3.94**	**0.10**	0.58	0.15	1.90	**0.24**	**0.12**
		MixN-GARCH-LOG	3.89**	4.25**	0.17	1.05	1.26	0.29	0.36
Conditional Coverage, LR_{CC}	1%	Normal-GARCH	7.54**	12.58***	11.84***	5.89	23.80***	0.40	1.65
		Normal-ASYM-GARCH	11.47***	16.19***	16.73***	18.06***	22.31***	0.66	2.58
		Normal-GJR-GARCH	8.44**	12.58***	9.68**	10.74***	22.31***	0.40	3.14
		Normal-EGARCH	13.74***	19.44***	25.37***	16.73***	26.59***	1.06	5.17
		MixN-GARCH	1.20	2.08	2.08	**0.81**	5.06	0.40	**0.54**
		MixN-GARCH-ASYM	**0.77**	6.82*	8.44**	3.14	5.95	**0.28**	0.99
		MixN-GARCH-GJR	1.29	5.97	3.14	1.65	**3.88**	0.40	0.64
		MixN-GARCH-LIK	1.20	**1.29**	**1.29**	1.89	5.95	0.40	0.55
		MixN-GARCH-LOG	2.77	6.82*	5.17	0.99	8.04**	0.66	0.62
	5%	Normal-GARCH	9.38**	2.25	1.38	**1.77**	2.27	0.84	4.91
		Normal-ASYM-GARCH	14.97***	4.53	3.64	7.38**	3.10	0.48	4.45
		Normal-GJR-GARCH	11.54***	2.31	**0.87**	3.92	3.10	0.84	3.85
		Normal-EGARCH	21.95***	12.17***	2.00	4.88	3.21	3.11	7.59**
		MixN-GARCH	6.48*	1.01	2.73	3.12	3.10	0.29	4.46
		MixN-GARCH-ASYM	11.04***	7.55**	3.56	9.91**	1.83	0.64	3.92
		MixN-GARCH-GJR	9.27**	10.42**	3.56	6.92*	1.85	**0.25**	**0.68**
		MixN-GARCH-LIK	**4.66**	**0.26**	0.92	2.81	3.10	**0.25**	1.35
		MixN-GARCH-LOG	12.32***	8.85**	3.12	7.22*	**1.78**	0.29	13.67***

Table 6: Test statistics at the 1%- and 5%-VaR level, $LR_{CC} = LR_{UC} + LR_{IND}$, as described in Christoffersen (1998) for all models under study. Entries in boldface denote the best outcomes. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Results are based on the same out-of-sample exercise as in Table 3.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Likelihood							
MixN-GARCH-LIK	-6031.16	-5188.23	-5073.16	-6277.30	-2330.67	**-2209.86**	-6349.75
MixN-GARCH-LIKW	**-6027.02**	**-5180.88**	**-5067.12**	**-6276.77**	**-2330.67**	-2210.78	**-6347.09**
tab:bic							
MixN-GARCH-LIK	**12136.05**	**10450.18**	**10220.04**	**12628.33**	**4731.03**	**4489.40**	**12773.23**
MixN-GARCH-LIKW	12144.14	10451.86	10224.33	12643.65	4746.51	4506.72	12784.28
Anderson-Darling							
MixN-GARCH-LIK	**0.84**	**0.57**	**0.95**	1.11	**0.66**	**0.61**	**1.26**
MixN-GARCH-LIKW	1.00	0.97	1.22	**1.11**	0.66	0.61	1.32
Cramér-von Mises							
MixN-GARCH-LIK	**0.11**	**0.07**	**0.14**	0.19	**0.07**	**0.10**	**0.15**
MixN-GARCH-LIKW	0.13	0.11	0.17	**0.17**	0.07	0.10	0.15
Kolmogorov-Smirnov (test statistics are scaled up by factor 100)							
MixN-GARCH-LIK	**1.65**	**1.23**	**1.76**	1.90	**1.76**	2.37	1.97
MixN-GARCH-LIKW	1.71	1.43	1.87	**1.81**	1.82	**2.35**	**1.84**
Ljung-Box ($m = 20$ lags)							
MixN-GARCH-LIK	28.26	**35.72****	**32.71***	**17.45**	18.65	**18.94**	**32.09***
MixN-GARCH-LIKW	**27.54**	36.25**	33.15*	17.60	**18.59**	18.98	33.70*
Jarque-Bera							
MixN-GARCH-LIK	25.57***	**40.92*****	**33.76*****	**3.76**	**15.94*****	**4.48**	**17.99*****
MixN-GARCH-LIKW	**20.90*****	84.62***	36.05***	3.96	16.00***	4.50	30.45***
Shapiro-Wilk (test statistic ν is transformed by $1000(1-\nu)$)							
MixN-GARCH-LIK	2.93***	**3.34*****	**3.31*****	1.72**	**4.33*****	**1.29**	**2.96*****
MixN-GARCH-LIKW	**2.60*****	5.15***	3.41***	**1.69****	4.34***	1.29	3.93***
1% VaR							
MixN-GARCH-LIK	**1.02**	**1.16**	**1.16**	0.78	1.68**	**1.09**	**0.95**
MixN-GARCH-LIKW	1.03	1.24	1.30*	**0.92**	**1.64****	1.09	0.89
5% VaR							
MixN-GARCH-LIK	**5.37**	**4.84**	**5.25**	**5.72***	**5.66**	5.03	5.48
MixN-GARCH-LIKW	5.66*	5.25	5.31	5.88**	5.72	**5.00**	**5.44**
RMSE up to 1% VaR							
MixN-GARCH-LIK	0.05	**0.18**	**0.16**	0.13	**0.31**	**0.13**	**0.06**
MixN-GARCH-LIKW	**0.04**	0.24	0.20	**0.08**	0.31	0.13	0.09
RMSE up to 5% VaR							
MixN-GARCH-LIK	**0.14**	**0.13**	**0.22**	**0.29**	**0.53**	**0.20**	**0.21**
MixN-GARCH-LIKW	0.20	0.33	0.36	0.39	0.54	0.20	0.25
Unconditional Coverage, LR_{UC}							
MixN-GARCH-LIK	**0.03**	**0.56**	**0.56**	1.56	5.16**	**0.07**	**0.05**
MixN-GARCH-LIKW	**0.03**	1.26	2.21	**0.17**	**4.14****	**0.07**	0.39
Independence, LR_{IND}							
MixN-GARCH-LIK	**1.17**	**0.72**	**0.72**	**0.32**	**0.79**	**0.33**	0.50
MixN-GARCH-LIKW	**1.17**	0.82	0.92	0.46	0.92	**0.33**	**0.43**
Conditional Coverage, LR_{CC}							
MixN-GARCH-LIK	**1.20**	**1.29**	**1.29**	1.89	5.95	**0.40**	**0.55**
MixN-GARCH-LIKW	**1.20**	2.08	3.14	**0.64**	**5.06**	**0.40**	0.81

Table 7: Results as in Table 1–6 but for MixN-GARCH-LIK and MixN-GARCH-LIKW. Entries in boldface denote the best outcomes. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Results are based on the same out-of-sample exercise as in Table 3.